

Box 1: Epi goes nano	184
A tougher code	183

Epigenome: mapping in motion

Monya Baker

As high-throughput techniques accelerate mapping of epigenetic marks, researchers are racing to find the biological meaning of these marks.

Thanks to the Human Genome Project, researchers worldwide can search a database to see what a gene 'says'. In just a few years, researchers may also be able to look up when a gene is 'read'. Or, rather, they will be able to pull up the epigenome, the set of chemical modifications to DNA and DNA-spooling proteins that coordinate how cells use genes.

Even with international enthusiasm, coordination and funding, epigenome mapping will be a long, complicated slog. Although an individual's genome sequence varies little from cell to cell, each of the 200 or so human cell types has its own epigenome. Epigenomes also change during development and in response to the environment. Cancer, aging and even behavioral disorders are all associated with epigenetic lesions. "The epigenome space is so much larger than the genome space," says John Stamatoyannopoulos of the University of Washington in Seattle, who heads one of four epigenome mapping centers funded by the US National Institutes of Health. "The measurement space is absolutely gigantic. No single technology is going to penetrate this with anything approaching completeness," he says.

Epigenetic marks are heritable modifications that alter gene expression without

affecting the DNA sequence. In particular, DNA is modified by enzymes that place methyl groups on cytosines. Histones, the protein complexes that package DNA in the nucleus, are modified by an army of enzymes that attach and remove a variety of chemical groups to and from particular amino acids.

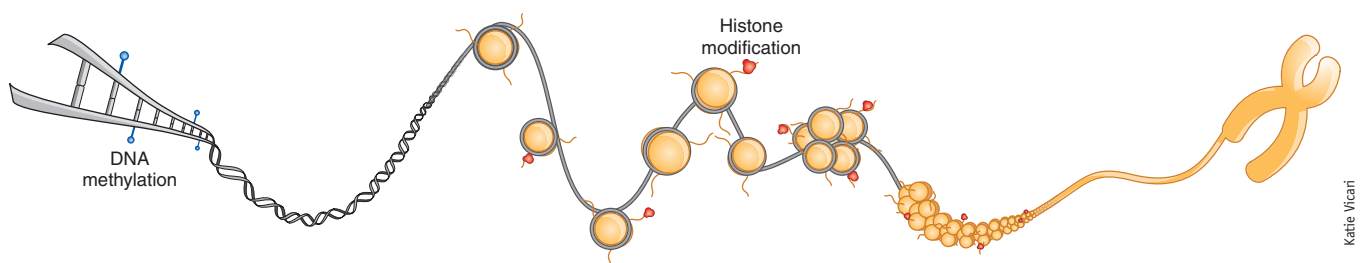
Mapping DNA methylation patterns is conceptually similar to the Human Genome Project. A tiny fraction of the billion-plus cytosine residues are marked with a methyl group; the vast majority of these marks are found on cytosines that precede guanines and are called CpGs. These methylation marks are, for the most part, faithfully copied whenever DNA replicates. Genes with highly methylated promoters are generally not expressed. In the past, most studies of human methylation looked at one or a few genes at a time and looked for methylation of the promoter. More recently, however, researchers have been able to broaden their studies. "Now you have the option to go from studying a single locus to studying all CpGs in the genome," says Alex Meissner of the Broad Institute.

Balancing coverage, costs and cohorts

When studying DNA methylation, researchers make trade-offs between the

time, money and cell numbers required to conduct experiments, completeness of coverage and the number of individuals or cell types that can be studied. In some cases, researchers want to know whether the frequency of methylation in regions of the genome differs between different cell types or patient populations. In other studies, the goal is to know the methylation status of individual cytosines in those regions.

The most complete study¹ so far was published in the fall of 2009. Researchers led by Bing Ren of the University of California, San Diego and Joseph Ecker of the Salk Institute profiled DNA methylation in human embryonic stem cells and fetal fibroblasts. Not only did they examine the 27 million CpG locations on the 23 pairs of human chromosomes, they assessed the status of over 90% of the billion-plus cytosines. As in perhaps most methylation studies, the researchers used a technique called sodium bisulfite conversion, in which a chemical treatment is used to convert unmethylated cytosines to uracil while leaving methylated cytosines unchanged. They then sequenced the converted DNA and compared it to an unconverted sequence to find positions of the methylated cytosines. Their study was



Epigenetic marks consist of chemical modifications to DNA and the histones that package it.

Katie Vrcari

comprehensive but expensive. Ecker says the work cost about \$100,000 per methylome at the time the work was completed. With Illumina's new HiSeq machine, he estimates the cost would come down to about \$20,000. In contrast, the most comprehensive microarray, the Illumina Infinium methylation assay, can be used to assess about 27,000 CpG dinucleotides and costs around \$225.

Right now, researchers who want to examine methylation in many genes use arrays, but whereas arrays are easy to use and (currently) cheaper than sequencing, they lack the resolution of sequencing,

says Joseph Costello of the University of California, San Francisco, another of the National Institutes of Health epigenome mapping centers. For example, arrays cannot be used to map methylation in retroposons and other repeat elements in the genome that, says Costello, "may turn out to be more than repeat elements." And although arrays can accurately indicate the presence of methylated cytosines in regions of 500 base pairs or so, they can leave those cytosines' exact location ambiguous.

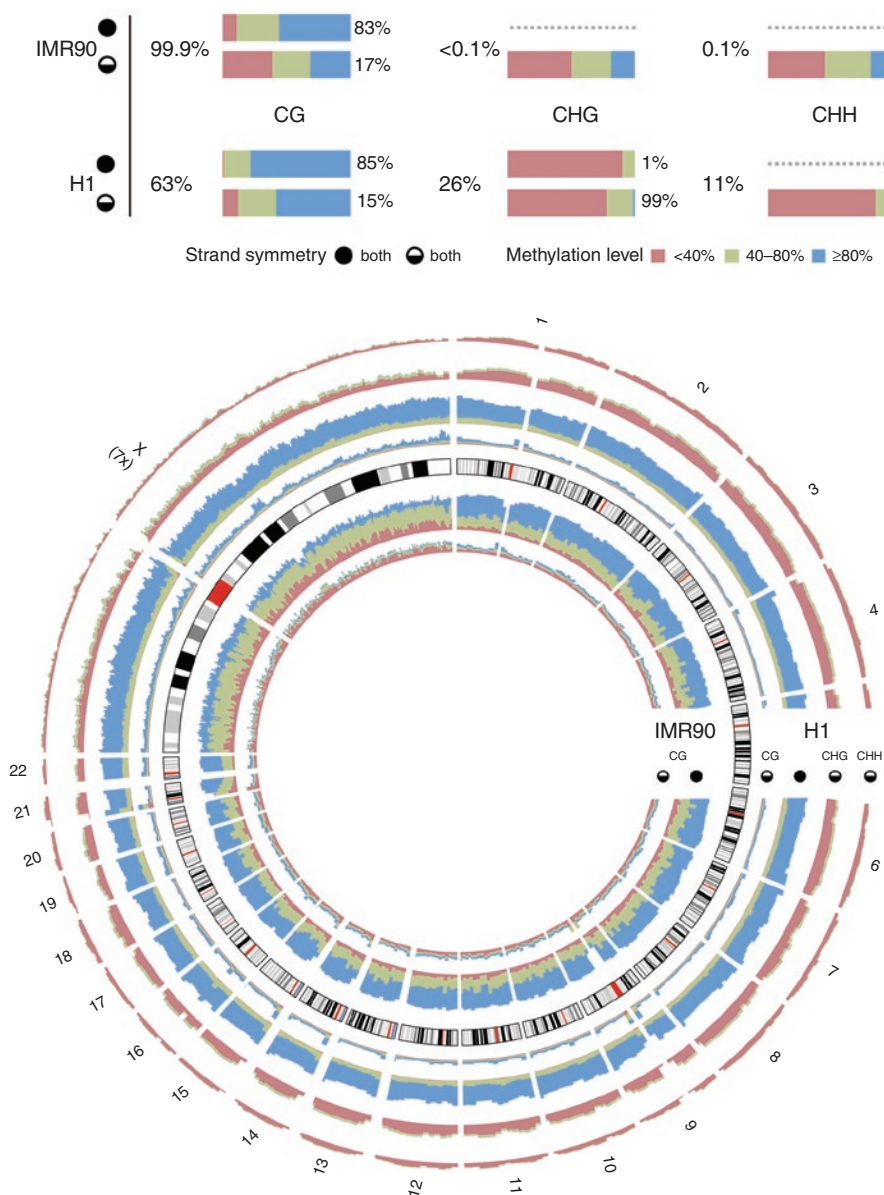
Perhaps as many as 80% of CpGs will be "boring" or unchanging between cell types, says Meissner. That understanding

could eventually help researchers design arrays or sequencing procedures to focus only on the interesting, changing bits, but, he says, "since we don't know them, we can't limit ourselves." This presents a conundrum for scientists like Andrew Feinberg of Johns Hopkins University. To understand the basis of common diseases, researchers need data from hundreds, maybe even thousands, of individuals, he says. "If you want to find something that's really new and important for human disease, you can't afford to do whole-genome sequencing, not on lots of samples."

A variety of array- and sequencing-based techniques are being used to help bring the epigenome down to size. Meissner has developed an approach called reduced-representation bisulfite sequencing, in which genomic DNA is cleaved by a restriction enzyme to enrich for sequences containing CpGs. The enzyme-cleaved fragments are then converted, amplified and sequenced. Meissner and colleagues recently reported using the technique on human clinical samples to obtain methylation data on over a million unique CpGs per sample². Examined regions included all sorts of genomic features: CpG islands containing dense collections of cytosines, untranslated regions, promoters, enhancers and others. "It's about as close as you can get to genome-wide without actually sequencing the whole genome," he says.

Other techniques also depend on enrichment. For example, methyl-DNA immunoprecipitation (meDIP) uses an antibody that latches onto methylated cytosines; captured DNA fragments can then be analyzed by sequencing or array analyses. Kun Zhang of the University of California, San Diego has developed so-called padlock probes that can be used to selectively amplify high-interest regions containing CpGs or other features of interest³. His team is currently synthesizing padlock probes that will cover over a half-million CpG sites that, from computational and empirical studies, are expected to be differentially methylated in various cell types.

Members of the four NIH epigenome mapping centers have recently completed studies comparing the different sequencing-based methylation methods side by side. "There is a lot of agreement in terms of what's methylated and what's not," says Costello. The biggest differences are in the information that techniques do



The first complete maps of human methylomes found surprising differences¹. In lung fetal fibroblasts (IMR90) methylation is observed only at cytosines adjacent to guanine; in human embryonic stem cells (H1), other cytosines are methylated. (Figure reproduced from ref. 5.)

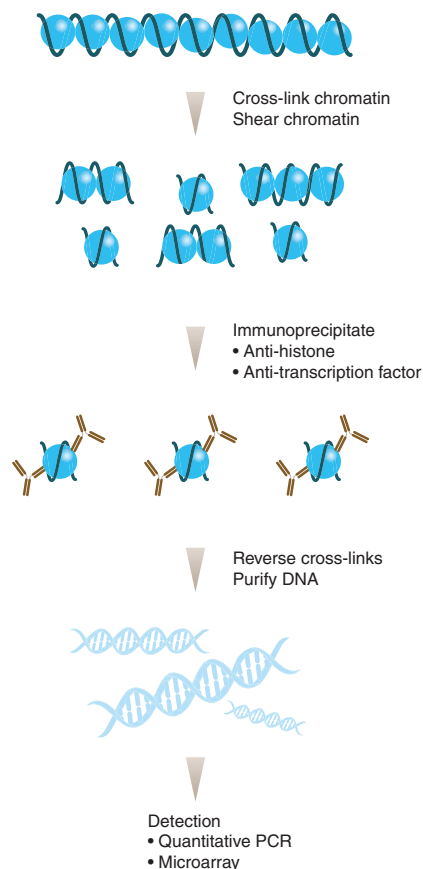
TECHNOLOGY FEATURE

not provide. Put simply, better resolution means worse coverage, except for the radically more expensive whole-genome shotgun technique. And, of course, the more expensive or extensive experiments are, the fewer samples can be examined and compared.

Also, there is concern that enrichment techniques miss interesting regions of the genome. In a recent study of human colon cancer, Feinberg and colleagues designed a DNA array in which sequences were arranged by the density of CpGs and found that more than three-quarters of differentially methylated regions were not within CpG-dense islands but near them⁴. They are now working with Zhang to develop padlock probes that target this region. The goal, Feinberg says, is to advance both science and discovery techniques. “We don’t know when high-throughput sequencing will come on board,” he says. “The methods are always evolving, and there are other things we can do now.”

The technology to sequence entire methylomes could also reveal unexpected complications, says Stephan Beck of University College London. In fact, the first published human methylome may have revealed one such complication. There are about 56 million CpG sites on the 46 human chromosomes; even broad enrichment techniques examine a tiny fraction of these. However, the whole-genome methylome study found that cytosines outside CpGs are often methylated in pluripotent stem cells^{1,5}. Such sites do not seem to be important in differentiated cells, says Beck, but if they are, researchers will have even more work ahead of them than they had planned. Two separate studies found that a considerable number of modified cytosines are not methylated but hydroxymethylated, and the dominant bisulfite sequencing techniques cannot distinguish between these two modifications^{6,7}. It is unclear still whether the modifications matter for biological function, Beck says, but it could mean a clear devaluation for a technique that is largely considered the gold standard for mapping methylation.

Beck is enthusiastic about techniques that might bypass this problem by detecting modified bases directly on single DNA molecules (Box 1), but he is certainly not one to advocate waiting for an optimal technology: he and colleagues launched a human epigenome project in 2000, even



Chromatin immunoprecipitation followed by DNA microarray or sequencing is used to discover what parts of the genome are associated with various histone modifications.

before the draft sequence of the human genome was available. Now, ten years later, international scientists and funding bodies have met together as the International Human Epigenome Consortium to coordinate efforts toward the same goal. The time is still right, he says. “We have to identify those tasks that are doable now, that are fundable now and then build up the program in the best possible way.”

A tougher code

If the human methylome, with its manifold variations, is harder to survey than the human genome, the maps for chromatin modifications will be more complicated still. Chromatin consists mainly of DNA wrapped around histones. Histones are protein octamers, each with two copies of four histone proteins, and these provide sites for over 100 post-translational modifications. Certain histone modifications are consistently associated with active or inactive genes, but these epigenetic signatures are far from simple.

BOX 1 EPI GOES NANO

Methylcytosines are sometimes called the fifth base, but they cannot be detected by standard sequencing techniques. Finding DNA methylation spots precisely usually means converting unmethylated cytosines to uracils, amplifying the converted DNA and comparing resultant sequences to a 'standard' sequence.

So it is no wonder that researchers are clamoring for ways to detect methylated cytosines directly, ideally conducting sequencing from a single DNA molecule. Not only would such techniques be faster and less susceptible to contamination, but they could work for cells available in only minute quantities, such as laser-dissected clinical samples or blastomeres from an early embryo. In October 2009, Pacific Biosciences received a \$1.2 million grant specifically to develop technology to detect DNA methylation. Oxford Nanopore Technologies, which raised 17.4 million pounds (\$28 million) in February 2010, has already described how one such device might work. A DNA nuclease is attached to a nanopore; as the nuclease cleaves a single-stranded DNA molecule, different nucleotides disrupt the flow of an electrical current to a different extent, enabling identification of dozens of base pairs per minute. Though the commercial device is still under development, company founder and Oxford University professor Hagan Bayley has shown that nanopores can distinguish all four nucleotides plus methylcytosine and believes that researchers will find this nanopore approach particularly suited to studying methylomes. "What they'd normally do is look for an averaged picture of all the methylation sites in all the cells. This will give you a picture of the whole genome and how it varies from cell to cell."

Another nanodevice deciphers chromatin modifications. At Cornell University, Paul Soloway and Harold Craighead are working on techniques to label specific chromatin modifications with different fluorescent tags; these labeled samples are made to flow through a nanotube toward a panel of lasers, which excite fluorophores, allowing researchers to detect modifications and sort them in a technique similar to flow cytometry.

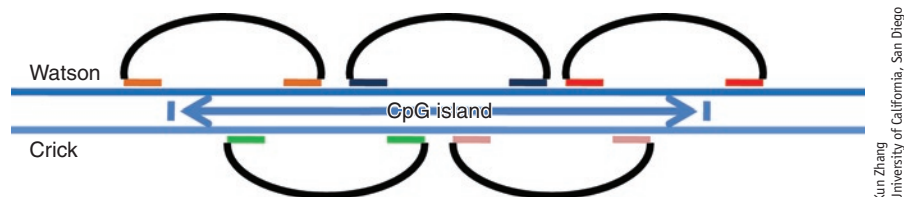
"There are so many different modifications; you don't know which one will be relevant to the biology that you are assessing," says Mathieu Lupien of Dartmouth University.

Despite the plethora of histone modifications, they are generally identified using the same general techniques. First, antibodies specific to a particular modification are added to fragmented chromatin, a technique known as chromatin immunoprecipitation, or ChIP. Then the associated DNA is analyzed either with microarrays or, increasingly commonly, by sequencing.

The NIH Epigenome Roadmap aims to catalog the locations of a core set of six well-studied histone modifications across

the genome in a hundred or so cell types. Another project will assess the presence of some 50 modifications in fewer cells. Yet another project will look for unrecognized types of modifications. Both of these exploratory projects can nominate additional histone marks to become part of the core set for more complete cataloging. (In addition to DNA methylation, other projects are exploring chromatin accessibility and small RNAs.)

Tracking combinations of histone marks will also be important, says David Allis of The Rockefeller University. Marks associated with conflicting outcomes can co-occur on the same histone, and hundreds, if not thousands, of unique combinations of modifications can be identified



Padlock probes allow selective amplification of over 100,000 CpG locations.

Kun Zhang
University of California, San Diego

within an individual cell type. For these kinds of studies, immunoprecipitation assays become tricky, he says, because different antibodies must be used sequentially. Another technique on the horizon for looking at genome-wide chromatin modifications is mass spectrometry.

But even researchers going after single, robust epigenetic marks are ending up with breathtaking results, says Allis, describing a recent study in cocaine-addicted mice. In January 2010, researchers at The Rockefeller University and Mount Sinai School of Medicine pinned addiction memory to a specific histone modification in certain brain cells, then tinkered with the enzymes responsible for the modifications. This affected the animals' preference for cocaine⁸. "When people start to look at these modifications and take it [follow-up experiments]

through to their biology and show that it makes a difference, that's pretty huge."

In fact, Life Technologies and Millipore both report seeing increased interest in learning how to correlate gene activity with both DNA methylation and histone modification. Entry of these new users into the market has caused a proliferation of kits carefully packed with controls and error-reducing workflows. Researchers studying stem cells and neuroscience are particularly keen, says Vasiliki Anest of Life Technologies. "Other researchers outside the [epigenetics] community are starting to see value in epigenetics and looking at how they can pose new questions." Just five years ago, researchers tended to be highly focused in terms of what aspects of the epigenome they wanted to investigate, says Sallie Cassel of Millipore. "As time has progressed they are not so

discrete anymore, people who are looking at [DNA] methylation may be looking at histones. People who are looking at RNAs may also be looking at DNA methylation. These groups are melting into a whole new melting pot called epigenetics."

1. Lister, R. *et al. Nature* **462**, 315–322 (2009).
2. Hongchang, G. *et al. Nat. Methods* **7**, 133–136 (2010).
3. Deng, J. *et al. Nat. Biotechnol.* **27**, 353–360 (2009).
4. Irizarry, R.A. *et al. Nat. Genet.* **41**, 178–186 (2009).
5. Costello, J.F., Krywinski, M. & Marra, M.A. *Nat. Biotechnol.* **27**, 1130–1132 (2009).
6. Kriaucionis, S. & Heintz, N. *Science* **324**, 929–930 (2009).
7. Tahiliani, M. *et al. Science* **324**, 930–935 (2009).
8. Maze, I. *et al. Science* **327**, 213–216 (2010).

Monya Baker is technology editor for *Nature* and *Nature Methods* (m.baker@us.nature.com).



TECHNOLOGY FEATURE

SUPPLIERS GUIDE: COMPANIES OFFERING EPIGENETICS SERVICES AND PRODUCTS

Company	Web address
454 Life Sciences (Roche)	http://www.454.com/
Active Motif	http://www.activemotif.com/
Affymetrix (USB)	http://www.affymetrix.com/
Agilent Technologies	http://www.agilent.com/
Applied Biosystems	http://www.appliedbiosystems.com/
Arrayit Corporation	http://www.arrayit.com/
Beckman Coulter	http://www.beckmancoulter.com/
BioChain	http://www.biotech.com/
Bruker Daltonics	http://www.bdal.com/
Clontech	http://www.clontech.com/
Diagenode	http://www.diagenode.com/
EpiGenie	http://www.epigenie.com/
Epigentek	http://www.epigentek.com/
Eurofins MWG Operon	http://www.operon.com/
Exiqon	http://www.exiqon.com/
Expression Analysis	http://www.expressionanalysis.com/
febit	http://www.febit.com/
GE Healthcare	http://www.gehealthcare.com/
Genisphere	http://www.genisphere.com/
GenScript	http://www.genscript.com/
Geospiza	http://www.geospiza.com/
Helicos BioSciences	http://www.helicosbio.com/
Human Genetic Signatures	http://www.geneticsignatures.com/
Illumina	http://www.illumina.com/
Intelligent Bio-Systems	http://www.intelligentbiosystems.com/
LC Sciences	http://www.lcsciences.com/
Life Technologies	http://www.lifetechnologies.com/
Millipore	http://www.millipore.com/
NABsys	http://www.nabsys.com/
Nanogen	http://www.nanogen.com/
New England Biolabs	http://www.neb.com/
Open Biosystems	http://www.openbiosystems.com/
Oxford Nanopore Technologies	http://www.nanoporetech.com/
Pacific Biosciences	http://www.pacificbiosciences.com/
Perkin Elmer	http://www.perkinelmer.com/
Phalanx Biotech Group	http://www.phalanxbiotech.com/
Promega	http://www.promega.com/
QIAGEN	http://www1.qiagen.com/
R&D Systems	http://www.rndsystems.com/
RainDance Technologies	http://www.raindancetechnologies.com/
Roche Applied Science	http://www.roche-applied-science.com/
Roche Nimblegen	http://www.nimblegen.com/
SA Biosciences	http://www.sabiosciences.com/
Santa Cruz Biotechnology, Inc	http://www.scbt.com/
Sequenom	http://www.sequenom.com/
SeqWright	http://www.seqwright.com/
SigmaAldrich	http://www.sigmaaldrich.com/
Zymo Research	http://www.zymoresearch.com/